Quaternary International 282 (2012) 137–144

Contents lists available at SciVerse ScienceDirect

Quaternary International

journal homepage: www.elsevier.com/locate/quaint

Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques: A case study of the Xiangxi River basin, China

Xingzhong Wang [a,b], Qinghua Cai [a,*], Lin Ye [a,b], Xiaodong Qu [a,b]

[a] State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, PR China
[b] Graduate University of Chinese Academy of Science, Beijing 100039, PR China

ARTICLE INFO

Article history:
Available online 10 May 2012

ABSTRACT

The analysis and interpretation the spatiotemporal patterns of river water quality are a critical element in the assessment, restoration, and protection of local and region water quality. In this case study, multivariate statistical techniques, including cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA), had been integrated to evaluate and interpret spatiotemporal variations of water quality in Xiangxi River, with a 5-years (2002–2006) continual monitoring data (14 parameters at 12 sites). Hierarchical cluster analysis revealed all sites could be grouped into three clusters representing different levels of pollution: relatively less polluted upper catchments sites (US), medium polluted Middle catchments sites (MS), and highly polluted lower catchments sites (LS). Factor analysis/principal component analysis was used to explore the most important factors determining the spatiotemporal dynamics of water quality in Xiangxi River. Varifactors obtained from the factor analysis indicated the parameters responsible for water quality variation were mainly related to soluble salts (natural), point source pollution of phosphorus and non-point pollution of nitrogen (anthropogenic). Discriminant analysis provided an important data reduction as it uses six parameters (TN, SiO_2, hardness, Ca^{2+}, WT and pH), affording 70.5% correct assignations in temporal analysis, and two parameters (NO_3–N and Alk), affording 55.9% correct assignations in spatial analysis, of three different regions in the basin. The low correct assignation in spatial analysis was related to the anthropogenic influence. This study suggested that multivariate statistical techniques are useful tools for identification of important water quality monitoring sites parameters and design of a monitoring network for the effective management of water resources.

1. Introduction

Rivers have always been the most important freshwater resources for human consumption, agricultural needs, and industrial and recreational purposes (Razmkhah et al., 2010; Varol et al., 2011). However, many rivers/streams in the developing countries are heavily polluted due to anthropogenic activities (Jonnalagadda and Mgere, 2001), especially in China (Jaehnig and Cai, 2010). Pollution of surface water with chemicals and excess nutrients is of great environmental concern worldwide (Ouyang, 2005; Koklu et al., 2010). The excess concentrations of chemicals and biologically available nutrients can lead to diverse problems such as toxic algal blooms, loss of oxygen, fish kills, loss of biodiversity and loss of aquatic plant beds (Voutsa et al., 2001). The degradation of water quality due to these contaminants has resulted in altered species composition and decreased the health of aquatic communities within the river basin (Ouyang et al., 2006). With an increased understanding of the important of drinking water quality to public health and raw water quality to aquatic life, there is a great need to assess water quality (Ouyang, 2005).

The evaluation of water quality in most countries has become a critical issue in recent years; especially because of concerns that freshwater will be a scarce resource in the future (Varol et al., 2011). However, there is a certain difficulty to interpret a huge and complex data matrix comprised of a large number of physico-chemical parameters form long-term monitoring programs (Bengraine and Marthaba, 2003; Singh et al., 2004; Koklu et al., 2010). Surface waters are most vulnerable to pollution due to their easy accessibility for disposal of wastewaters (Singh et al., 2004). The water quality of river at any point can reflected several major influences, including the lithology of the basin, atmospheric inputs, climatic conditions and anthropogenic inputs (Bricker and Jones, 1995), as well as interactions between several factors (Bengraine and Marthaba, 2003).

* Corresponding author.
E-mail address: qhcai@ihb.ac.cn (Q. Cai).

1040-6182/$ – see front matter © 2012 Elsevier Ltd and INQUA. All rights reserved.
doi:10.1016/j.quaint.2012.05.015

The multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA) has widely been used to interpret water quality data in the identification of possible factors/source that influence water systems and offer a valuable tool for reliable management of water resource (Singh et al., 2004; Zeng and Rasmussen, 2005; Ouyang et al., 2006; ; Shrestha and Kazama, 2007; Hussain et al., 2008; Kazi et al., 2009; Razmkhah et al., 2010; Varol et al., 2011). The multivariate statistical treatment of data also has been applied to characterize and evaluate surface and freshwater quality and it is useful for verifying temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality (Helena et al., 2000; Singh et al., 2004, 2005; Shrestha and Kazama, 2007).

As the largest tributary of the Three Gorges Reservoir (TGR) in Hubei Province, the Xiangxi River can strongly influence water quality of the TGR (Zhou et al., 2008; Wu et al., 2009). Therefore water quality monitoring and assessment is very important for the management of TGR area. In the present study, a large data matrix, obtained during a 5-year (2002–2006) monitoring program, is subjected to different multivariate statistical techniques to: (1) obtain information about the similarities and differences between sampling sites, (2) identify water quality variable responsible for spatial and temporal variations in stream water quality, and (3) determine the influence of sources (natural and anthropogenic) on water quality parameters of the Xiangxi River.

2. Materials and methods

2.1. Study area

Xiangxi River, located in central China, is the largest tributary of the Three Gorges Reservoir (TGR) in Hubei Province. The Xiangxi River originates from Mount Shennongjia (3150 m a.s.l., the highest mountain in central China), and it discharges into the Yangtze River. It has three main tributaries: the Jiuchong, Gufu and Gaolan River (Fig. 1). Xiangxi River watershed has an area of 3099 km^2 and a natural fall of 1540 m from its headwaters to its confluence with the Yangtze River (Jiang et al., 2005; Tang et al., 2006). The average annual precipitation within this watershed is 988 mm (Cai et al., 2010; Li et al., 2010).

Approximately 70.9% of this catchment is covered by forests, 6.5% is farmland and 5.3% is water area, 4.4% is wasteland whilst the remainder is residential land and transportation (Xu et al., 2010). Forests are mainly distributed on hillsides the slope upper regions, and have obvious changes in dominant tree species along altitudinal gradients (Jiang et al., 2005). The main agricultural crops are rice and wheat. Urban areas are mainly distributed in the middle and lower regions of the basin near the river bank. Soils are derived mainly from limestone soils in the upper regions and brown and yellow-brown soils in the lowlands (Hörmann et al., 2009).

2.2. Analytical procedure

The data sets of 12 water quality monitoring sites comprising 14 physical–chemical parameters monitored monthly over 5-years (2002–2006) (Fig. 1). A hydrolab Minisonde (Hach Environmental, Loveland, Colorado) was used to measure in situ variables that included pH, conductivity (Cond), and water temperature (WT) at each site.

Surface water samples also were collected in two 380 ml cleaned plastic containers to measure chemical variables according to the standard methods (Huang, 1999; Cai, 2007) in the lab, including total nitrogen (TN), ammonium (NH_4–N), nitrate (NO_3–N), total phosphorus (TP), orthophosphate (PO_4–P), hardness, calcium (Ca^{2+}), chloride (Cl^-), alkalinity (Alk), silicon (SiO_2) and chemical oxygen demand (COD).

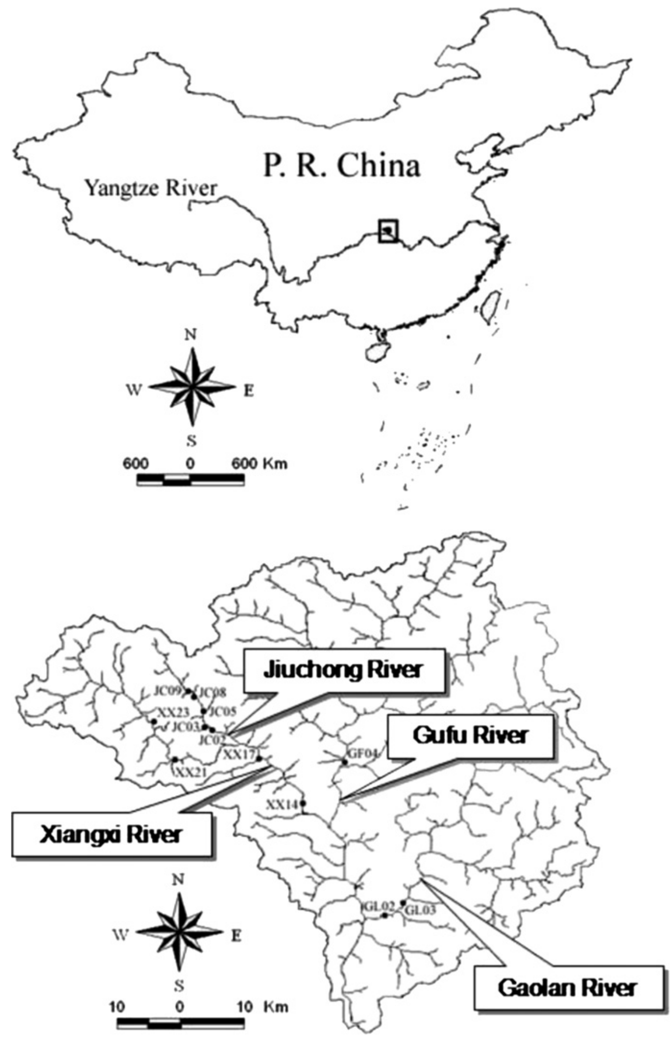

Fig. 1. Location of the Xiangxi River watershed in China and the distribution of sampling sites.

2.3. Data treatment

The Kolmogorov–Smirnov (K–S) statistics were used to test the goodness-of-fit of the data to normal distribution. According to the K–S test, measured water-quality parameters show non-normal distribution. Spearman's rank correlation coefficient was used to assess the correlation structure between non-normal distributed water quality parameters (Wunderlin et al., 2001). Therefore, the temporal variations of the stream water-quality parameters were first evaluated through season–parameter correlation matrix using the Spearman non-parametric (Spearman's R) in this study. The water-quality parameters were grouped in four seasons: spring (March–May), summer (June–August), autumn (September–November) and winter (December–February). Each season was assigned to a numerical value in the data file (spring = 1; summer = 2; autumn = 3 and winter = 4), which as a variable corresponding to the season was correlated (pair by pair) with all the measured parameters.

Stream water quality data sets information was interpret by integrating four multivariate analysis: cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA) (Wunderlin et al., 2001; Singh et al., 2004;

Shrestha and Kazama, 2007; Hussain et al., 2008; Kazi et al., 2009; Razmkhah et al., 2010; Varol et al., 2011). CA and PCA/FA were applied on experimental data standardized through z-scale transformation in order to avoid misclassification due to wide differences in data dimensionality (Liu et al., 2003; Simeonov et al., 2003). DA was applied to raw data (Singh et al., 2004; Shrestha and Kazama, 2007). All mathematical and statistical computations were made using Microsoft Office Excel 2003, SPSS 16.0 and STATISTICA 6.

3. Results and discussion

3.1. Spatial similarity and site grouping

Cluster analysis yielded a dendrogram (Fig. 2), grouping all 12 sampling sites of the basin into three statistically significant clusters at $D_{link}/D_{max} \times 100 < 60$. The clustering procedure generated three groups are very meaningful, as the sites in these groups have similar characteristic features and natural background source types. Cluster 1 (GF04, GL03 and GL02), cluster 2 (JC02, JC03, XX14 and XX17) and cluster 3 (JC08, JC09, XX21 and XX23) correspond to high pollution lower catchments region (LS), moderate pollution middle catchments region (MS) and low pollution upper catchments region (US), respectively. The sites of Cluster 1 are located at the downstream of the urban area, may affected by point source discharges of sewage as well as the pollution from phosphorus industry. Cluster 2 and Cluster 3 sites are respectively located in an agricultural dominated region and forest dominated region. Many previous studies demonstrated the pattern of good water quality in forest areas and high concentrations of nitrogen in agricultural areas (Allan, 2004). It indicates that CA technique is useful to make reliable classification of steam water quality in the whole region and help to develop future spatial sampling strategy in an optimal manner.

3.2. Data structure determination and source identification

Kaiser–Meyer–Olkin (KMO) and Bartlett's test were needed to examine the suitability of the data for PCA/FA before analysis. KMO is a measure of sampling adequacy that indicates the proportion of variance which is common variance. High value (close to 1) generally indicates that PCA/FA may be useful. In this study, KMO = 0.494. Bartlett's test indicates whether correlation matrix in is an identity matrix or not, which indicates that variables are

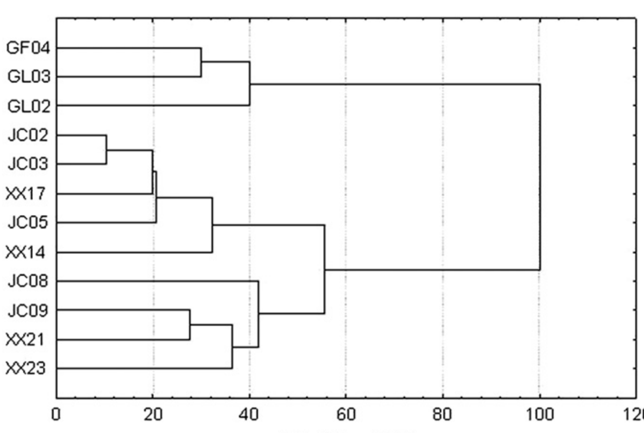

Fig. 2. Dendrogram showing clustering of sampling sites on the Xiangxi river according to water quality characteristic.

Table 1
Loadings of experimental variables (14) on significant principal components for spring, summer, autumn and winter data sets.

Variables	VF1	VF2	VF3	VF4	VF5	VF6
Spring (six significant principal components)						
NH_4–N	−0.013	−0.018	0.037	−0.001	0.101	**0.910**
NO_3–N	0.331	**0.821**	−0.236	0.139	−0.154	0.053
TN	0.126	**0.936**	0.157	−0.048	−0.076	−0.006
PO_4–P	−0.228	0.472	0.076	−0.259	*0.570*	−0.239
TP	0.182	−0.107	−0.064	**0.859**	0.013	0.107
SiO_2	−0.223	0.393	0.240	*0.678*	−0.221	−0.328
Alk	*0.704*	0.280	0.070	−0.142	0.118	−0.225
Hardness	*0.705*	0.158	0.295	0.059	0.097	0.246
Ca	0.064	0.212	*0.725*	−0.185	−0.108	0.395
Cl	0.078	−0.112	**0.860**	0.077	0.042	−0.143
COD	−0.435	0.029	−0.139	0.414	0.153	−0.064
WT	0.059	0.098	0.193	0.094	**−0.851**	−0.151
Cond	**0.867**	0.008	−0.149	0.189	−0.137	−0.008
pH	0.135	−0.211	0.191	0.237	0.540	0.051
Eigenvalue	2.23	2.15	1.62	1.62	1.51	1.32
%Total variance	19.81	13.78	12.53	11.10	9.45	8.02
Cumulative %variance	19.81	33.60	46.13	57.22	66.67	74.70
Summer (five significant principal components)						
NH_4–N	−0.055	**0.809**	−0.050	−0.198	0.177	
NO_3–N	**0.863**	0.035	0.231	0.120	0.082	
TN	**0.841**	−0.044	−0.067	−0.009	−0.240	
PO_4–P	0.231	*0.583*	−0.190	0.170	−0.290	
TP	0.034	*0.748*	−0.111	−0.079	−0.152	
SiO_2	*0.649*	−0.211	0.154	−0.141	*0.536*	
Alk	−0.053	−0.214	**0.787**	−0.196	−0.124	
Hardness	0.134	−0.024	**0.844**	0.153	0.088	
Ca	−0.284	0.005	0.172	*0.706*	0.329	
Cl	−0.013	0.121	−0.044	**−0.773**	0.098	
COD	−0.360	**0.766**	0.073	−0.071	0.185	
WT	−0.068	0.035	−0.021	0.044	**0.853**	
Cond	0.465	0.020	*0.616*	0.082	0.466	
pH	0.309	−0.081	−0.202	*0.627*	−0.045	
Eigenvalue	2.48	2.26	1.92	1.67	1.61	
%Total variance	22.31	14.76	14.04	11.56	8.36	
Cumulative %variance	22.31	37.07	51.10	62.66	71.02	
Autumn (six significant principal components)						
NH_4–N	−0.118	−0.064	*0.597*	−0.323	0.484	0.074
NO_3–N	**0.907**	−0.060	0.049	0.131	−0.081	0.159
TN	**0.916**	−0.052	−0.154	−0.014	0.011	0.082
PO_4–P	−0.021	−0.127	0.147	0.463	*0.521*	−0.246
TP	0.041	0.053	0.010	0.010	**0.770**	0.203
SiO_2	0.083	0.206	−0.046	**0.888**	0.053	0.070
Alk	0.336	−0.018	0.434	−0.063	*−0.578*	0.191
Hardness	0.105	**0.868**	−0.067	0.049	−0.177	−0.081
Ca	0.342	**−0.797**	−0.019	−0.002	−0.033	0.173
Cl	*0.566*	−0.137	0.331	−0.267	−0.092	−0.479
COD	0.193	−0.088	−0.027	−0.001	0.056	**0.868**
WT	0.034	−0.029	**0.757**	0.042	−0.035	−0.150
Cond	−0.135	0.002	*0.684*	*0.598*	−0.077	0.083
pH	−0.025	*0.739*	−0.022	0.124	0.132	0.142
Eigenvalue	2.30	2.03	1.75	1.58	1.51	1.25
%Total variance	19.38	14.26	13.34	11.57	8.35	7.57
Cumulative %variance	19.38	33.64	46.99	58.55	66.90	74.47
Winter (five significant principal components)						
NH_4–N	**0.794**	0.062	0.006	−0.359	−0.204	
NO_3–N	0.065	0.149	*0.531*	*0.640*	0.173	
TN	**0.756**	0.133	0.171	0.285	0.149	
PO_4–P	−0.118	**0.943**	0.088	0.066	−0.060	
TP	0.240	**0.929**	0.024	−0.011	0.024	
SiO_2	**0.803**	0.039	−0.048	0.336	0.125	
Alk	−0.210	0.070	**0.897**	0.007	−0.039	
Hardness	0.365	−0.028	**0.805**	−0.001	0.054	
Ca	0.298	0.320	0.017	−0.237	*0.520*	
Cl	−0.150	−0.352	0.106	−0.016	0.369	
COD	0.002	0.103	0.178	**−0.762**	−0.014	
WT	0.263	−0.176	0.343	−0.070	*0.510*	
Cond	0.400	0.169	0.215	*0.690*	−0.062	
pH	−0.116	−0.033	−0.169	0.332	*0.681*	
Eigenvalue	2.45	2.10	2.01	1.97	1.25	
%Total variance	23.12	14.49	12.46	11.94	7.84	
Cumulative %variance	23.12	37.61	50.08	62.02	69.86	

Bold and italic values indicate strong and moderate loadings, respectively.

Table 2
Classification functions (Eq. (3)) for discriminant analysis of temporal variation in water quality of the Xiangxi river basin.

Parameters	Standard mode				Forward stepwise mode				Backward stepwise mode			
	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter
	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]
NH_4–N	24.562	19.906	16.465	18.832	25.396	20.765	17.241	19.609				
NO_3–N	−3.914	−5.575	−4.430	−4.838	−4.336	−6.009	−4.822	−5.231				
TN	1.682	5.422	2.233	2.113	1.867	5.612	2.404	2.285	2.059	5.038	2.366	2.053
PO_4–P	10.022	7.491	6.195	3.851	10.520	8.003	6.658	4.315				
TP	12.553	16.474	20.159	16.848	9.834	13.674	17.633	14.316				
SiO_2	0.384	−0.525	−0.083	0.334	0.421	−0.487	−0.049	0.368	−0.165	−1.049	−0.707	−0.216
Alk	0.260	0.253	0.266	0.275	0.269	0.262	0.274	0.283				
Hardness	−0.027	−0.033	0.329	0.106	−0.016	−0.021	0.339	0.116	0.411	0.367	0.755	0.556
Ca	0.299	0.183	0.356	0.437	0.299	0.183	0.356	0.437	0.386	0.260	0.435	0.519
Cl	0.739	0.761	0.687	0.688								
COD	1.105	1.299	1.126	1.266	1.034	1.225	1.060	1.200				
WT	1.314	2.214	1.661	0.843	1.408	2.311	1.748	0.931	1.416	2.301	1.738	0.974
Cond	−0.005	−0.004	−0.010	−0.004	−0.008	−0.007	−0.013	−0.007				
pH	24.935	25.082	23.810	25.435	24.832	24.976	23.714	25.340	23.222	23.392	22.142	23.709
Constant	−142.01	−152.77	−139.82	−148.61	−140.70	−151.39	−138.69	−147.47	−116.64	−127.74	−114.92	−121.43

[a] Discriminant function coefficient for spring, summer, autumn and winter seasons correspond to w_{ij} as defined in Eq. (1).

unrelated. The significance level which is 0 in this study (less than 0.05) indicates that there are significant relationships among variables, and suitable for PCA/FA. The principal component analysis results were compared for the whole normalized data, normalized data sets separately for the three regions and normalized data sets separately for the four seasons. The PCA result was best performed on normalized data sets separately for the four seasons. PCA of the four data sets yielded six PCs for spring and autumn and five PCs for summer and winter with eigenvalues >1, explaining 74.70%, 74.47%, 71.02% and 69.86% of the total variance in respective water quality data sets. Equal numbers of VFs were obtained for four seasons through FA performed on the PCs. PCA/FA were widely used to assess spatial and temporal variation in water quality (Singh et al., 2004). Liu et al. (2003) classified the factor loadings as 'strong', 'moderate' and 'weak' corresponding to absolute loading values of >0.75, 0.75–0.50, and 0.50–0.30, respectively.

For the data set pertaining to spring, among six VFs (Table 1), VF1 explaining 19.81% of total variance, has strong positive loading on Cond and moderate positive loading on Alk and hardness, which can be interpreted as a mineral component of the stream water quality. These minerals were likely from dissolution of limestone and gypsum soils (Vega et al., 1998). VF2, explaining 13.78% variance, has strong positive loading on NO_3–N and TN. The excess nitrogen in Xiangxi River was mainly from town sewage and agriculture cultivation (Li et al., 2007; Ye et al., 2009). VF3 (12.53% of total variance) has strong positive loading on Cl^- and moderate positive loading on Ca^{2+}. VF4 (11.10% of total variance), has strong positive loading on TP and moderate positive loading on SiO_2. VF5 (9.45% of total variance) has strong negative loading on WT and moderate positive loading on PO_4–P. VF3, VF4 and VF5 represent mineral and nutrition component. Water temperature in natural can strong impact on dissolution of limestone and gypsum soils, also have strong relationship with the dissolution and inversion of phosphorus. The phosphorus reserves in the Xiangxi River basin is among the top three in China, reach 357 million tons (Li et al., 2008). The phosphorite and phosphate plant were the point sources polluted the Xiangxi River. VF6 explaining 8.02% of the total variance, has strong positive loading on NH_4–N, reflected the nitrogen pollution of town sewage.

For the data set representing the summer, among five significant VFs (Table 1), VF1 explains 22.31% of the total variance, has strong

Table 3
Classification matrix for discriminant analysis of temporal variation in water quality of the Xiangxi river basin.

Monitoring seasons	% Correct	Season assigned by DA			
		Spring	Summer	Autumn	Winter
Standard DA mode					
Spring	64.2	70	3	15	21
Summer	83.9	5	94	13	0
Autumn	56.0	14	21	56	9
Winter	87.9	11	0	5	116
Total	74.2	100	118	89	146
Forward stepwise DA mode					
Spring	64.2	70	3	14	22
Summer	84.2	6	96	12	0
Autumn	57.0	14	20	57	9
Winter	85.6	14	0	5	113
Total	73.8	104	119	88	144
Backward stepwise DA mode					
Spring	56.4	66	4	17	30
Summer	80.5	10	99	14	0
Autumn	51.5	20	19	52	10
Winter	88.1	11	0	5	118
Total	70.5	107	122	88	158

positive loading on NO_3–N and TN, moderate positive loading on SiO_2. VF2 (14.76% of total variance) has strong positive loading on NH_4–N and COD and moderate positive loading on PO_4–P and TP. VF1 and VF2 represent the anthropogenic pollution sources and can be explained that the high precipitation in summer make the no-point pollution more serious. VF3 explaining 14.04% of the total variance, has strong positive loading on Alk and hardness and moderate positive loading on Cond. VF4 (11.56% of total variance) has strong negative loading on Cl^- and moderate positive loading on Ca^{2+}. VF5 (8.36% of total variance), has strong positive loading on WT and moderate positive loading on SiO_2. VF3, VF4 and VF5 represent the mineral component and water temperature. This can be explained that the high precipitation in summer make more dissolution of limestone and gypsum soils.

For the data set representing the autumn, among total six significant VFs (Table 1), VF1 explaining 19.38% of the total variance, has strong positive loading on NO_3–N and TN and moderate positive loading on Cl^-. This VF reflects the pollution of nitrogen, which may be a result of agriculture harvest. VF2 explaining 14.26% of the total variance, has strong positive loading on pH. VF3 (13.34% of total variance) has strong positive loading on WT, and moderate positive loading on Cond. VF4 (11.57% of total variance) has strong positive loading on SiO_2 and moderate positive loading on Cond. VF2, VF3 and VF4 reflect the dissolution of limestone and gypsum soils and the change of water temperature. VF5 explains 8.35% of the total variance, has strong positive loading on TP, moderate positive loading on PO_4–P and moderate negative loading on Alk. VF6 (7.57% of total variance) has strong positive loading on COD. VF5 and VF6 represent influences from point source, such as phosphorite and phosphate plant.

Lastly, for the data set pertaining to water quality in winter, among five VFs (Table 1). VF1 explaining 23.12% of the total variance, has strong positive loading on NH_4–H, TN and SiO_2. VF2 (14.49% of total variance) has strong positive loading on PO_4–P and TP. VF3 (12.46% of total variance) has strong positive loading on Alk and hardness, and moderate positive loading on NO_3–N. VF4

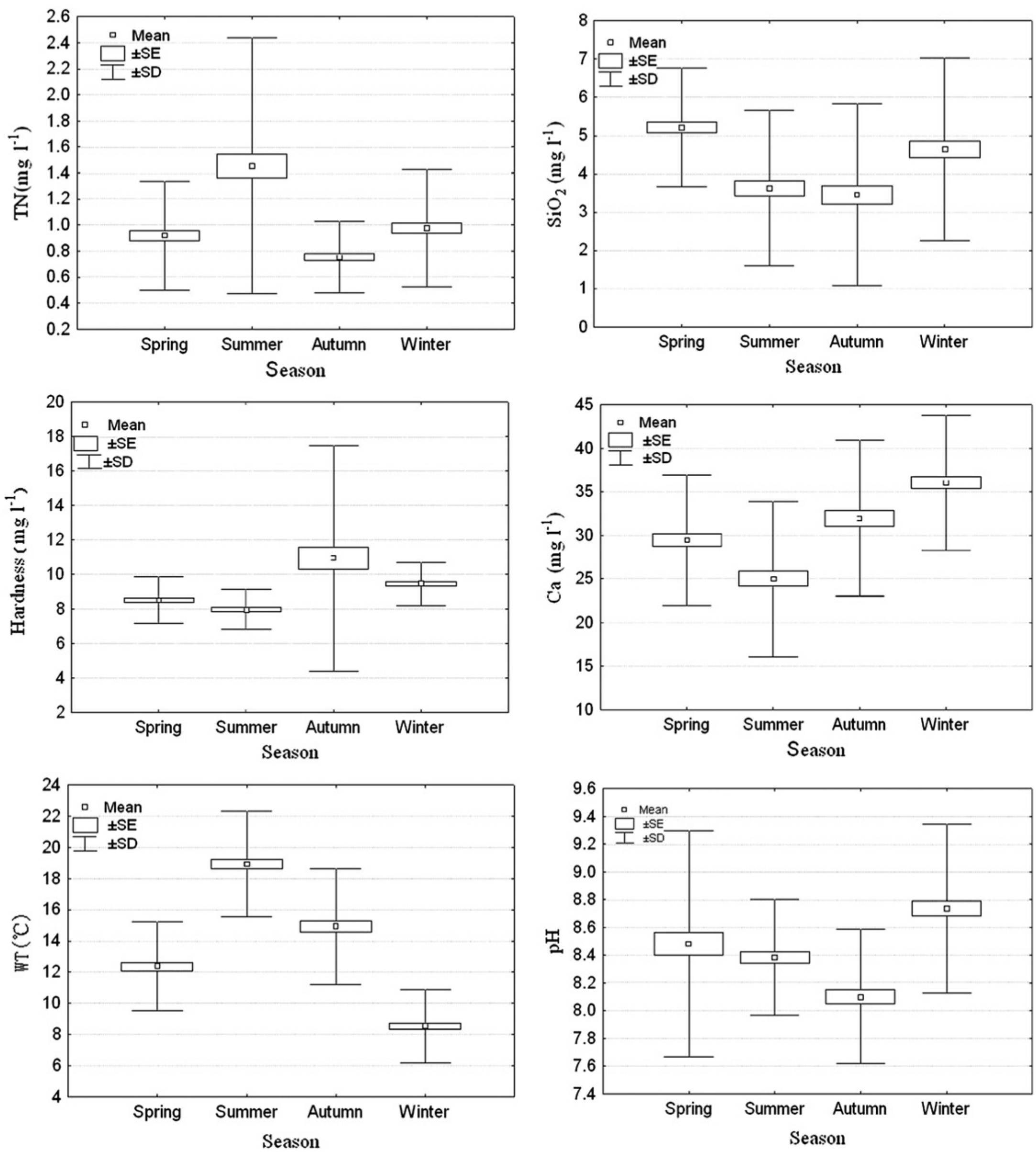

Fig. 3. Temporal variations: TN, SiO_2, hardness, Ca, WT and pH in surface water quality of the Xiangxi river basin.

Table 4
Classification functions (Eq. (3)) for discriminant analysis of spatial variation in water quality of the Xiangxi river basin.

Parameters	Standard mode			Forward stepwise mode			Backward stepwise mode		
	LS[b]	MS[c]	US[d]	LS[b]	MS[c]	US[d]	LS[b]	MS[c]	US[d]
	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]	Coefficient[a]
NH_4–N	23.701	24.009	21.646						
NO_3–N	4.885	1.856	0.224	5.401	2.304	0.791	8.132	5.310	4.042
TN	−0.907	−0.665	−0.294	0.307	0.620	0.925			
PO_4–P	5.472	6.404	8.050	−0.613	0.351	2.019			
TP	22.615	22.968	19.172	19.912	20.077	15.854			
SiO_2	0.962	0.801	0.609	1.667	1.503	1.320			
Alk	0.336	0.318	0.294	0.293	0.273	0.250	0.264	0.246	0.226
Hardness	0.492	0.457	0.389	0.676	0.654	0.583			
Ca	0.389	0.370	0.377						
Cl	1.044	0.960	0.835	0.844	0.761	0.632			
COD	1.528	1.288	1.203	1.488	1.249	1.144			
WT	1.617	1.574	1.510	0.769	0.729	0.661			
Cond	−0.006	−0.010	−0.009	0.010	0.007	0.007			
pH	22.988	23.169	23.235						
Constant	−154.54	−146.29	−140.40	−46.78	−37.64	−31.10	−26.34	−20.61	−17.49

[a] Discriminant function coefficient for spring, summer, autumn and winter seasons correspond to w_{ij} as defined in Eq. (1).
[b] Lower catchments includes sites (GF04, GL03 and GL02).
[c] Middle catchments includes sites (JC02, JC03, XX14 and XX17).
[d] Upper catchments includes sites (JC08, JC09, XX21 and XX23).

(11.94% of total variance) has strong negative loading on COD and moderate positive loading on NO_3–N and Cond. VF1 and VF3 reflect the mineral component and nitrogen pollution in the stream water quality. VF2 and VF4 reflect the nutrition pollution of phosphorus and nitrogen. VF5 explaining 7.84% of the total variance, has moderate positive loading on Ca^{2+}, WT and pH, reflect the dissolution of limestone and gypsum soils.

3.3. Temporal and spatial variations in stream water quality

The temporal variations in stream water quality parameters were evaluated through a season–parameter correlation matrix. The result showed that 10 parameters were found to be significantly ($p < 0.01$) correlated with season, while 4 parameters were found to be not significantly ($p > 0.05$) correlated with season. The all season–parameter correlation coefficients were generally low. Among these, WT exhibited highest correlation coefficient (Spearman's $R = -0.40$) followed by Ca^{2+} (Spearman's $R = 0.33$) and hardness (Spearman's $R = 0.32$). The season-correlated parameters can be taken as representing the major source of temporal variances in water quality (Singh et al., 2004; Shrestha and Kazama, 2007). Wide seasonal variations in water temperature and stream discharge can be attributed to the high seasonality in various water quality parameters (Singh et al., 2004; Shrestha and Kazama, 2007). The non-significant correlation of NH_4–N, TN, Cl^- and Cond with season indicates the contribution of anthropogenic sources in the catchment areas.

Table 5
Classification matrix for discriminant analysis of spatial variation in water quality of the Xiangxi river basin.

Monitoring regions	% Correct	Season assigned by DA		
		LS[a]	MS[b]	US[c]
Standard DA mode				
LS	41.2	42	53	7
MS	73.2	19	150	36
US	58.9	3	57	86
Total	61.4	64	260	129
Forward stepwise DA mode				
LS	41.7	48	60	7
MS	74.5	18	172	41
US	54.8	4	71	91
Total	60.7	70	303	139
Backward stepwise DA mode				
LS	31.6	43	91	2
MS	76.7	23	217	43
US	43.4	8	108	89
Total	55.9	74	416	134

[a] Lower catchments includes sites (GF04, GL03 and GL02).
[b] Middle catchments includes sites (JC02, JC03, XX14 and XX17).
[c] Upper catchments includes sites (JC08, JC09, XX21 and XX23).

Temporal variations in water quality were further evaluated through DA. Temporal DA was performed on raw data of four seasons (spring, summer, autumn and winter). Discriminant functions (DFs) and classification matrices (CMs) obtained from the standard, forward stepwise and backward stepwise mode are shown in Tables 2 and 3. The standard DA mode, constructed DFs including 14 parameters, yielded the corresponding CMs assigning 74.2% of the cases correctly (Tables 2 and 3). Both the standard and forward stepwise mode DFs using 13 and 6 discriminant variables, respectively, rendered the corresponding CMs assigning 74% cases correctly (Tables 2 and 3). However, in backward stepwise mode DA gave CMs with 70.5% correct assignations using only six discriminant parameters with a litter different match for each season compared with the forward stepwise mode. Thus, the temporal DA results suggested that TN, SiO_2, hardness, Ca^{2+}, WT and pH are the most significant parameters to discriminate between the four seasons, which means that these six parameters account for most of the expected temporal variations in the stream water quality.

As identified by DA (back stepwise mode), box and whisker plots of the selected parameters showing seasonal trends are given in Fig. 3. The variation of water temperature shows a clear-cut seasonal effect. The average concentration of TN is high in summer can be explained that high precipitation in summer make the anthropogenic non-point pollution in to stream water body serious. The average concentration of Ca^{2+} and hardness due to dissolution of limestone and gypsum soils were low in summer, suggested it has negative relationship with high precipitation. The average pH and SiO_2 concentration have the similar trend.

Spatial DA was performed with the same raw data set comprising 14 parameters after grouping into three regions of LS, MS and US obtained through CA. DFs and CMs obtained from modes are shown in Tables 4 and 5. Similarly to temporal DA, the standard

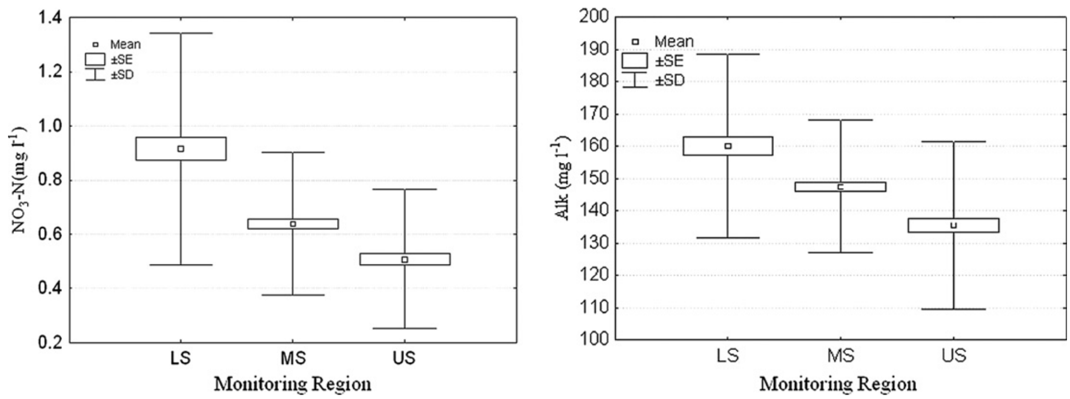

Fig. 4. Spatial variations: NO_3–N and Alk in surface water quality of the Xiangxi river basin.

mode, constructed DFs including 14 parameters, yielded the corresponding CMs assigning 61.4% of the cases correctly. The forward and backward stepwise mode used 11 and 2 discriminant variables, yielded the corresponding CMs assigning 60.7% and 55.9% cases correctly, respectively. Back stepwise mode shows that NO_3–N and Alk are the discriminating parameters in space.

According to the spatial DA (Back stepwise mode), box and whisker plots of discriminating parameters were constructed to evaluate different patterns associated with spatial variations in stream water quality (Fig. 4). The concentration of NO_3–N was increased from upstream region to downstream region. Ye et al. (2009) suggested that the agriculture sub-watershed has high concentration of NO_3–N and Alk and the forest dominated region has low concentration of most nutrient variables in Xiangxi basin. The case correctly of DA was low due to the high deviation within groups suggested that there have point pollution source of nitrogen. The high deviations within groups also have relationship with temporal pattern of the anthropogenic non-point pollution of nitrogen in Xiangxi River basin. There are many small hydropower station in Xiangxi River basin (Wu et al., 2010), but it did not affect any water-chemistry variables (Wu et al., 2009). The average Alk concentration has the similar trend to the average NO_3–N concentration.

4. Conclusions

With the concerns of river water quality in recent years, it was required to develop wide range multivariate statistical techniques to analyze and interpret underlying water quality information. In this study, multivariate statistical techniques, including Cluster Analysis (CA), Principal Component Analysis (PCA), Factor Analysis (FA) and Discriminant Analysis (DA), had been integrated for the assessment, restoration, and protection of local and region water quality. CA grouped 12 sampling sites into three clusters of similar water quality characteristics, which may be a result of different land use. This study suggested that good water quality in forested area and high concentrations of nitrogen in agricultural areas. Extracted grouping information can be used in reducing sampling sites without losing much information. Although the PCA/FA did not resulted in a significant data reduction at four seasons, the VFs obtained from the PCs suggested that the parameters responsible for water quality variations are mainly related to the dilution of salt (natural), the point source pollution of phosphorus and the pollution of nitrogen (agriculture cultivation in spring, harvest in autumn and town sewage). Discriminant analysis provided an important data reduction. For four seasons of the basin, it yielded good data reduction, as it used six parameters (TN, SiO_2, hardness, Ca^{2+}, WT and pH), affording 70.5% correct assignations. For three different sampling regions, it yielded two parameters (NO_3–N and Alk), affording 55.9% correct assignations. The low correct assignations in spatial analysis suggested the anthropogenic influence of agriculture cultivation in spring and harvest in autumn. Thus, future watershed management activities should consider the effect of land use as well as the point pollution from the phosphorus industry.

Acknowledgements

The authors gratefully acknowledge Daofeng Li, Tao Tang, Shuchan Zhou, Meiling Shao, Naicheng Wu, Fengqing Li, Xinghuan Jia and Shugui Duan for their assistance with the field sampling. The authors also thank Ruiqiu Liu and Yaoyang Xu for their assistance in measurement of water quality. This work was funded by the National Natural Science Foundation of China (30330140, 40671197), and State key Laboratory FEBL Research Grant (2011FB202).

References

Allan, J.D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology Evolution and Systematics 35, 257–284.

Bengraine, K., Marthaba, T.F., 2003. Using Principal component analysis to monitor spatial and temporal changes in water quality. Journal of Hazardous Materials B100, 179–195.

Bricker, O.P., Jones, B.F., 1995. Main factors affecting the composition of natural waters. In: Salbu, B., Steinnes, E. (Eds.), Trace Elements in Natural Waters. CRC Press, Boca Raton, FL, pp. 1–5.

Cai, Q. (Ed.), 2007. Protocols for Standard Observation and Measurement in Aquatic Ecosystems. Chinese Environmental Science Press, Beijing (in Chinese).

Cai, Q., Liu, M., He, Y., Zeng, X., Jiang, T. (Eds.), 2010. China Climate Change Impact Report: Yangtze Three Gorges Area. China Meteorological Press, Beijing (in Chinese).

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernández, J.M., Fernández, L., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Research 34, 807–816.

Huang, X., 1999. Survey, Observation and Analysis of Lake Ecology. China Standards Press, Beijing (in Chinese).

Hussain, M., Ahmed, S.M., Abderrahman, W., 2008. Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. Journal of Environmental Management 86, 279–307.

Hörmann, G., Köplin, N., Cai, Q., Fohrer, N., 2009. Using a simple model as a tool to parameterise the SWAT model of the Xiangxi River in China. Quaternary International 208, 116–120.

Jaehnig, S.C., Cai, Q., 2010. River water quality assessment in selected Yangtze tributaries: background and method development. Journal of Earth Science 21 (6), 876–881.

Jiang, M., Deng, H., Cai, Q., Wu, G., 2005. Species richness in a riparian plant community along the banks of the Xiangxi River, the Three Gorges region. The International Journal of Sustainable Development and World Ecology 12 (1), 60–67.

Jonnalagadda, S.B., Mgere, G., 2001. Water quality of the Odzi river in the eastern highlands of Zimbabwe. Water Research 35 (10), 2371–2376.

Kazi, T.G., Arain, M.B., Jamali, M.K., Jalbani, N., Afridi, H.I., Sarfraz, R.A., Baig, J.A., Shah, A.Q., 2009. Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. Ecotoxicology and Environmental Safety 72, 301–309.
Koklu, R., Sengorur, B., Topal, B., 2010. Water quality assessment using multivariate statistical methods – a case study: Melen River system (Turkey). Water Resource Management 24, 959–978.
Li, F., Ye, L., Liu, R., Cao, M., Cai, Q., 2007. Investigation on aquatic environmental factors in Xiangxi River watershed. Ecological Science 26 (3), 199–207 (in Chinese with English abstract).
Li, F., Cai, Q., Ye, L., 2010. Developing a benthic index of biological integrity and some relationship to environmental factors in the subtropical Xiangxi River, China. International Review of Hydrobiology 95 (2), 171–189.
Li, F., Ye, L., Liu, R., Cao, M., Cai, Q., 2008. Variations of the main nutrients input to Xiangxi Bay of the three Gorge Reservoir. Acta Ecologica Sinica 28 (5), 2073–2079 (in Chinese with English abstract).
Liu, C.W., Lin, K.H., Kuo, Y.M., 2003. Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Science of Total Environment 313, 77–89.
Ouyang, Y., Nkedi-kizza, P., Wu, Q.T., Shinde, D., Huang, C.H., 2006. Assessment of seasonal variations in surface water quality. Water Research 40, 3800–3810.
Ouyang, Y., 2005. Evaluation of river water quality monitoring station by principal component analysis. Water Research 39, 2621–2635.
Razmkhah, H., Abrishamchi, A., Torkian, A., 2010. Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: a case study on Jajrood River (Tehran, Iran). Journal of Environmental Management 91, 852–860.
Shrestha, S., Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. Environment Modelling & Software 22, 464–475.
Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M., Kouimtzis, T., 2003. Assessment of the surface water quality in Northern Greece. Water Research 37, 4119–4124.
Singh, K.P., Malik, A., Mohan, D., Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. Water Research 38, 3980–3992.
Singh, K.P., Malik, A., Sinha, S., 2005. Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques: a case study. Analytica Chimica Acta 538, 355–374.
Tang, T., Cai, Q., Liu, J., 2006. Using epilithic diatom communities to assess ecological condition of Xiangxi River system. Environment Monitoring and Assessment 112, 347–361.
Varol, M., Gökot, B., Bekleyen, A., Sen, B., 2011. Water quality assessment and apportionment of pollution sources of Tigris River (Tukey) using, multivariate statistical techniques – a case study. n/a. River Research and Applications 27. doi:10.1002/rra. 1553.
Vega, M., Pardo, R., Barrado, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research 32, 3581–3592.
Voutsa, D., Manoli, E., Samara, C., Sofoniou, M., Stratis, I., 2001. A study of surface water quality in Macedonia, Greece: speciation of nitrogen and phosphorus. Water Air and Soil Pollution 129, 13–32.
Wu, N., Tang, T., Fu, X., Jiang, W., Li, F., Zhou, S., Cai, Q., Fohrer, N., 2010. Impacts of cascade run-of-river dams on benthic diatoms in the Xiangxi River, China. Aquatic Science 72, 117–125.
Wu, N., Tang, T., Zhou, S., Jia, X., Li, D., Liu, R., Cai, Q., 2009. Changes in benthic algal communities following construction of a run-of-river dam. Journal of the North American Benthological Society 28, 69–79.
Wunderlin, D.A., Diaz, M.P., Ame, M.V., Pesce, S.F., Hued, A.C., Bistoni, M.A., 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba-Argentina). Water Research 35, 2881–2894.
Xu, H., Yaylor, R.G., Kingston, D.G., Jiang, T., Thompson, J.R., 2010. Hydrological modeling of River Xiangxi using SWAT2005: a comparison of model parameterizations using station and gridded meteorological observations. Quaternary International 226, 54–59.
Ye, L., Cai, Q., Liu, R., Cao, M., 2009. The influence of topography and land use on water quality of Xiangxi River in Three Gorges Reservoir region. Environmental Geology 58, 937–942.
Zeng, X., Rasmussen, T.C., 2005. Multivariate statistical characterization of water quality in lake Lanier, Georgia, USA. Journal of Environmental Quality 34, 1980–1991.
Zhou, S., Tang, T., Wu, N., Fu, X., Cai, Q., 2008. Impacts of a small dam on riverine zooplankton. International Review of Hydrobiology 93, 297–311.